Transcript: How Elite AI Agencies Build Prompts - The PDER Process

Video ID: i70LfKB6TqU

Extraction Date: 2025-04-03 02:14:56

**[00:00:00]** over the past two years my AI agency

**[00:00:01]** morning side AI has built AI systems for

**[00:00:03]** companies of all sizes from local

**[00:00:06]** martial arts gyms to publicly traded

**[00:00:08]** companies and even an NBA team and at

**[00:00:10]** the core of every one of our client

**[00:00:11]** projects are a handful of powerful AI

**[00:00:13]** prompts which in a few hundred carefully

**[00:00:16]** chosen words can automate hundreds or

**[00:00:18]** even thousands of hours of manual work

**[00:00:20]** so being able to write really really

**[00:00:21]** good prompts like these is a superpower

**[00:00:24]** these days regardless of what career

**[00:00:26]** you're in but the tricky thing is that

**[00:00:27]** writing these kinds of really good

**[00:00:29]** prompts usually takes a lot of time and

**[00:00:31]** effort realizing this at Morningside AI

**[00:00:33]** it became a major priority for us to

**[00:00:35]** figure out how to write these kind of

**[00:00:37]** elite production grade prompts for our

**[00:00:39]** client projects as quickly as possible

**[00:00:41]** and so over the past 2 years using the

**[00:00:43]** latest in AI prompting research and

**[00:00:45]** professional grade prompting software we

**[00:00:47]** created our own rapid prompt engineering

**[00:00:49]** system called the Peter process and in

**[00:00:51]** this video I'm going to be revealing our

**[00:00:52]** entire system including the exact prompt

**[00:00:54]** engineering tools and software we use

**[00:00:56]** and the process as well that we use at

**[00:00:58]** Morningside AI to craft hyper effective

**[00:01:00]** prompts in just minutes that are able to

**[00:01:02]** extract the most value from AI models

**[00:01:05]** possible without having to spend all day

**[00:01:06]** writing and testing those prompts so

**[00:01:08]** we're going to start quickly with an

**[00:01:09]** overview of the Peter framework and then

**[00:01:11]** diving into a more Hands-On walkthrough

**[00:01:13]** of exactly what it looks like when it's

**[00:01:14]** in action and we're basically going to

**[00:01:16]** be building a business grade prompt from

**[00:01:18]** scratch using this process so whether

**[00:01:20]** you're an employee who's just keen on

**[00:01:22]** automating some of the work you do or

**[00:01:24]** you're an AI agency owner wanting to

**[00:01:25]** deliver better client results and more

**[00:01:27]** consistently or if you're a business

**[00:01:29]** owner and you're looking to train up

**[00:01:30]** your own team on how to write prompts

**[00:01:32]** well and write them fast and be able to

**[00:01:34]** automate their own work then this video

**[00:01:35]** will show you how the professionals do

**[00:01:38]** it so first and foremost let's just talk

**[00:01:40]** about what makes prompt engineering so

**[00:01:41]** powerful right it's kind of overused and

**[00:01:43]** over talked about but when done right it

**[00:01:45]** is essentially like having a magic wand

**[00:01:47]** that can transform data from one form

**[00:01:49]** into another you're taking some kind of

**[00:01:51]** raw information and then turning it into

**[00:01:53]** something valuable just by writing some

**[00:01:55]** magic words in the middle it's pretty

**[00:01:57]** magical when you think about it but

**[00:01:58]** here's the reality that most people face

**[00:01:59]** right there's a massive difference

**[00:02:01]** between casual prompting that most

**[00:02:03]** people do and professional prompt

**[00:02:04]** engineering which is what I'm going to

**[00:02:05]** teach you how to do in this video so

**[00:02:07]** when you are just chatting to chat gbt

**[00:02:09]** you can go back and forth and you can

**[00:02:11]** kind of refine your requests by asking

**[00:02:13]** follow-up prompts and sending more more

**[00:02:14]** questions than nudg in the right

**[00:02:15]** directions until you get what you want

**[00:02:17]** so it may take a few tries but

**[00:02:18]** eventually you will get to the outcome

**[00:02:20]** you want but in the real world whether

**[00:02:22]** you're Building Systems for businesses

**[00:02:24]** like myself and the team at morning side

**[00:02:25]** AI or you're creating some client

**[00:02:27]** deliverables or you're automating entire

**[00:02:28]** workflows you need to be able to prompt

**[00:02:31]** at a level of what's called single shot

**[00:02:33]** prompting basically your prompt needs to

**[00:02:35]** be able to work perfectly the first time

**[00:02:37]** every time without any human

**[00:02:39]** intervention and kind of going back and

**[00:02:40]** forth and saying hey look could you

**[00:02:43]** could you do it without these words or

**[00:02:44]** in this kind of tone right you don't

**[00:02:46]** have that back and forth it needs to be

**[00:02:47]** able to be given data transform that how

**[00:02:50]** you want it and do that reliably at

**[00:02:51]** scale and this is where all of the value

**[00:02:53]** really is for prompt engineering as a

**[00:02:55]** skill but this is also where most people

**[00:02:56]** get stuck in this frustrating trade-off

**[00:02:58]** that I mentioned earlier on one hand you

**[00:03:00]** can either write prompts quickly that

**[00:03:01]** perform okay or you can spend hours

**[00:03:04]** crafting the ones that work flawlessly

**[00:03:06]** and when you're trying to build

**[00:03:07]** something valuable whether that's for

**[00:03:08]** your employer or for your clients or for

**[00:03:10]** your own business neither option is

**[00:03:11]** acceptable let me give you some quick

**[00:03:13]** examples of why this matters so say

**[00:03:15]** you're an employee right imagine you are

**[00:03:17]** automating a weekly report that normally

**[00:03:19]** takes you 3 hours so a mediocre prompt

**[00:03:22]** that you do quickly might get you maybe

**[00:03:24]** 70% of the way there there's still an

**[00:03:26]** hour or so of cleanup work that you have

**[00:03:27]** to do whereas a great prompt that you

**[00:03:29]** spend a lot of time on might do 95% of

**[00:03:32]** the work for you and pretty much save

**[00:03:33]** you that entire 3 hours and for AI

**[00:03:35]** automation agency owners when you are

**[00:03:37]** delivering to your clients the

**[00:03:39]** difference between a good and a great

**[00:03:40]** prompt can mean the difference between a

**[00:03:42]** one-time project and a long-term

**[00:03:44]** retainer and relationship with thousands

**[00:03:46]** and thousands of dollars per month or if

**[00:03:47]** you're a business owner training your

**[00:03:49]** team to write Elite prompts means that

**[00:03:51]** you can Implement AI across your entire

**[00:03:52]** organization with consistent and

**[00:03:54]** reliable results so what was in

**[00:03:55]** realizing the difference between these

**[00:03:56]** good and great prompts and how much

**[00:03:58]** extra leg work you can get the AI

**[00:04:00]** systems to do with the great prompts but

**[00:04:02]** also factoring in that these take a lot

**[00:04:03]** longer to write that we decided to

**[00:04:05]** create our own process at morning side

**[00:04:07]** AI to always get the best prompts that

**[00:04:08]** get the most out of AI models without

**[00:04:10]** having to spend all day writing and

**[00:04:12]** testing

**[00:04:13]** them so the Peter process is a

**[00:04:15]** systematic approach that allows us to

**[00:04:17]** create these high performing prompts in

**[00:04:18]** a fraction of the time that it would

**[00:04:19]** normally take us it's so simple that

**[00:04:21]** anyone can be trained on it so that they

**[00:04:23]** can use it in their own role for

**[00:04:24]** automating work so it's very helpful not

**[00:04:26]** just for our AI engineers and developers

**[00:04:28]** when they're doing client work it's it's

**[00:04:29]** helpful to teach our entire team so that

**[00:04:31]** they know how to create their own AI

**[00:04:33]** tools within their work and automate

**[00:04:34]** different parts of their own work that

**[00:04:35]** they would rather not be doing so it's a

**[00:04:37]** win-win across the whole organization

**[00:04:38]** when people are using this kind of

**[00:04:40]** process so P or PD stands for firstly p

**[00:04:43]** is plan which is clearly defining what

**[00:04:45]** you need the prompt to do before even

**[00:04:47]** writing a single word so this is

**[00:04:48]** something that many people chop up on

**[00:04:49]** and don't even do D is for draft and

**[00:04:52]** this is rapidly generating an initial

**[00:04:53]** prompt using an AI tool that we've

**[00:04:55]** created that basically rapidly applies

**[00:04:57]** the latest research back techniques when

**[00:04:59]** it comes to prompting then e is evaluate

**[00:05:02]** and this is using a professional prompt

**[00:05:03]** engineering software which I'll show you

**[00:05:04]** in a second in order to strategically

**[00:05:06]** test the prompt that againsts multiple

**[00:05:08]** expected inputs so it's rapid testing in

**[00:05:10]** a special prompt engineering environment

**[00:05:12]** and finally R is for refine which is

**[00:05:15]** looking at the responses that you get

**[00:05:16]** from the tests we do within software and

**[00:05:18]** then making targeted improvements and

**[00:05:20]** then repeating that cycle over and over

**[00:05:21]** again and what makes this process so

**[00:05:23]** powerful is that it combines three

**[00:05:24]** critical elements that most people Miss

**[00:05:26]** in their prompt engineering which is

**[00:05:28]** firstly using research back prompting

**[00:05:30]** techniques that are already baked into

**[00:05:31]** this process it's using custom AI tools

**[00:05:33]** that generate solid first drafts of the

**[00:05:35]** prompt for you saving you tons of time

**[00:05:37]** and thirdly using a professional prompt

**[00:05:39]** engineering IDE that most people have

**[00:05:41]** never heard of and the best part is that

**[00:05:42]** I'm going to be sharing all of these

**[00:05:43]** tools with you for free that we use at

**[00:05:45]** morning side so you're going to get

**[00:05:46]** access to our prompt planning worksheet

**[00:05:48]** our AI prompt generator tool and also a

**[00:05:50]** link to the professional prompt IDE that

**[00:05:52]** we use as well and so for the rest of

**[00:05:54]** this video to show you how this looks in

**[00:05:55]** action I'm going to be walking you

**[00:05:56]** through each step of the Peter process

**[00:05:58]** while actually building a production

**[00:06:00]** grade prompt from scratch so you're

**[00:06:01]** going to get to see exactly how we do it

**[00:06:02]** at morning side with nothing held back

**[00:06:04]** so let's start with the first component

**[00:06:05]** which is

**[00:06:07]** planning the planning phase is where

**[00:06:09]** most people go wrong before they even

**[00:06:10]** started really they just jump straight

**[00:06:12]** into writing a prompt without clearly

**[00:06:13]** defining what they needed to do so at

**[00:06:15]** morning side we start every prompt with

**[00:06:17]** a planning worksheet that answers some

**[00:06:18]** very key questions that you need to know

**[00:06:20]** before going into things questions like

**[00:06:22]** what are the inputs and what data is

**[00:06:23]** this prompt actually going to be

**[00:06:24]** processing what format does the output

**[00:06:27]** need to be in where can I find some high

**[00:06:29]** quality examples of input and output

**[00:06:30]** pairs which model should be used based

**[00:06:32]** on the budget and performance needs of

**[00:06:34]** this prompt and of this use case what

**[00:06:37]** are the specific requirements or

**[00:06:38]** constraints that have come to me from

**[00:06:39]** the client or from the the project as a

**[00:06:41]** whole so this planning phase typically

**[00:06:43]** takes 5 to 10 minutes but it can save

**[00:06:44]** hours and hours of frustration later so

**[00:06:46]** here on screen we have the prompt

**[00:06:47]** planning sheet which you guys are going

**[00:06:48]** to be able to get a copy of this is a

**[00:06:49]** resource directly out of my accelerator

**[00:06:51]** as with some of the other stuff I'm

**[00:06:52]** going to be sharing now before we can

**[00:06:53]** actually get into planning we need to

**[00:06:55]** know what kind of prompt we're going to

**[00:06:56]** be writing so in this video I've come up

**[00:06:58]** with a hypothetical uh scenario say you

**[00:07:00]** are an AI automation agency delivering a

**[00:07:02]** project for a client or maybe you're a

**[00:07:04]** business and you're doing this

**[00:07:05]** internally but we're going to be

**[00:07:06]** creating a prompt that is an email order

**[00:07:08]** responder um that can basically classify

**[00:07:10]** inbound emails to this fantasy uh

**[00:07:12]** e-commerce that I've come up with which

**[00:07:14]** is called Tech Gear Pro and they sell

**[00:07:15]** Tech products online and this prompt

**[00:07:17]** that we're about to write is going to be

**[00:07:18]** attached to their uh email inbox so

**[00:07:20]** every new email is going to be run

**[00:07:21]** through the prompt we write and the

**[00:07:23]** prompt is basically expected to be able

**[00:07:24]** to classify firstly so what kind of

**[00:07:27]** query is this is asking about order

**[00:07:29]** status are they asking for a return or a

**[00:07:31]** refund are they having technical issues

**[00:07:32]** and they need technical support or does

**[00:07:34]** it fall outside of these so we have four

**[00:07:36]** broad categories and that's the

**[00:07:37]** classification step and then within each

**[00:07:39]** of these we have a bit more detail and a

**[00:07:41]** bit more complexity CU I thought that

**[00:07:42]** was a bit too simple I want to give you

**[00:07:43]** guys a bit more of an advanced example

**[00:07:45]** so that you really see how you can apply

**[00:07:47]** this to difficult problems within

**[00:07:48]** businesses or even within your own maybe

**[00:07:50]** you're an employee and you do this stuff

**[00:07:51]** so if the email comes in and it

**[00:07:52]** classifies it as an order status inquiry

**[00:07:55]** if it's been 0 to 7 Days the process is

**[00:07:57]** basically to reassure them that hey look

**[00:07:59]** usually take 7 to 10 days um please just

**[00:08:01]** wait a few days you can track your order

**[00:08:02]** here at this ra but if it's over 8 days

**[00:08:05]** then this thing is going to go back and

**[00:08:06]** say hey look I'm going to be looking

**[00:08:08]** into this for you this should be here by

**[00:08:09]** now and that's basically buying the

**[00:08:11]** human teen time so in the real world um

**[00:08:13]** this is an autoresponder that deals with

**[00:08:15]** as much of the stuff up front as it can

**[00:08:16]** like hey your order should be on the way

**[00:08:18]** no worries just check this page but if

**[00:08:20]** it's gone over that then it can flag the

**[00:08:21]** human team and say hey look um give an

**[00:08:23]** instant response to the customer so that

**[00:08:25]** they're not sitting there getting all

**[00:08:26]** anti because they're not getting a

**[00:08:28]** response and basically buys the team

**[00:08:29]** time to look into it and then come in

**[00:08:31]** with another response the same thing

**[00:08:32]** happens with uh return and refund

**[00:08:34]** requests here obviously for most

**[00:08:36]** e-commerce stores they've got maybe

**[00:08:37]** 30-day returns policy so um part of the

**[00:08:39]** input data that we're going to be

**[00:08:40]** getting with these emails which we'll

**[00:08:42]** see in a second this will make a lot

**[00:08:43]** more sense in in a second but part of

**[00:08:45]** the data that we're going to be getting

**[00:08:46]** is also how many days since they Place

**[00:08:49]** their order so if it's been less than 30

**[00:08:50]** days since they purchased then the AI

**[00:08:52]** autoresponder can automatically give

**[00:08:53]** them the process for doing that return

**[00:08:55]** and getting the refund if it's been more

**[00:08:57]** then they sort of buy a better time for

**[00:08:58]** the team and say Hey look let me check

**[00:08:59]** if there's anything else I can do um and

**[00:09:01]** then the human team can step and same

**[00:09:02]** thing for technical support fairly

**[00:09:04]** straightforward it's trying to handle

**[00:09:05]** most of the situations up front but if

**[00:09:06]** it is a b more complex then it's buying

**[00:09:08]** time for the human so that's really the

**[00:09:10]** the value of the system it gets back to

**[00:09:12]** customers as quickly as possible answers

**[00:09:13]** as many of those easy to answer

**[00:09:15]** questions or just like routing them to

**[00:09:17]** the right resources as you can see with

**[00:09:18]** these links but in the case where the

**[00:09:20]** human does need a step and it's buying

**[00:09:21]** time and giving them a response while

**[00:09:23]** the humans come and solve it when they

**[00:09:24]** come onto their next shift and so this

**[00:09:25]** whole thing here is basically a a brief

**[00:09:27]** or a write up of what this project would

**[00:09:29]** be about and it's got all the context we

**[00:09:30]** need in order to be able to build a

**[00:09:32]** system to execute on this so now that we

**[00:09:33]** know what we're trying to do here we can

**[00:09:35]** go back to our planning sheet and fill

**[00:09:36]** this out quickly so what are the inputs

**[00:09:38]** if any inputs are going to be the uh

**[00:09:40]** customer email and the uh number of days

**[00:09:44]** since purchase so that's what we're

**[00:09:47]** expecting to be passed into this prompt

**[00:09:49]** and the sort of system as a whole once

**[00:09:50]** it's baked into some kind of workflow

**[00:09:51]** automation whereas the IP coming from a

**[00:09:54]** any in many cases The Prompt is needing

**[00:09:55]** to apply some kind of thinking or some

**[00:09:58]** kind of process and in this case the IP

**[00:10:00]** is coming from this document here which

**[00:10:01]** is all of this uh information so all of

**[00:10:03]** this is basically the IP in other cases

**[00:10:05]** where you're writing these prompts it

**[00:10:06]** may be for you're writing scripting or

**[00:10:09]** you're doing short FM content scripting

**[00:10:10]** or any kind of other content repurposing

**[00:10:12]** in this IP section you could have some

**[00:10:14]** Alexi you could have some previous

**[00:10:16]** examples of high performing scripts that

**[00:10:17]** you put out on your social media Etc but

**[00:10:19]** it's basically like what's the juice

**[00:10:21]** what's the meat what's actually uh

**[00:10:22]** telling it how to transform this data

**[00:10:24]** what are the instructions on how it

**[00:10:25]** should be doing its task so in this case

**[00:10:27]** where's IP coming from is um

**[00:10:30]** project brief what format does the

**[00:10:32]** output need to be in just a plain text

**[00:10:36]** email um respond so the output of this

**[00:10:39]** prompt is just going to be the email

**[00:10:40]** that the AI has written to respond to so

**[00:10:42]** that's not too complicated in a bunch of

**[00:10:44]** use cases for this kind of prompt

**[00:10:45]** engineering you'd want it to Output in

**[00:10:47]** Json so that you can do something with

**[00:10:48]** it uh via another step in a workflow

**[00:10:50]** automation or you want to extract some

**[00:10:52]** information and put into a database so

**[00:10:53]** the output format is important and how

**[00:10:55]** where can I get two to three high

**[00:10:56]** quality examples as you're going to see

**[00:10:58]** in the uh for prompt generator tool we

**[00:11:00]** do need input output examples in order

**[00:11:02]** to get the best possible prompt for my

**[00:11:04]** first draft generation in this case I've

**[00:11:05]** also included it on uh this write up so

**[00:11:08]** I'll give you guys access to all these

**[00:11:10]** resources it'll be in my free school

**[00:11:11]** Community you can just go onto School

**[00:11:13]** join for free it'll take a few minutes

**[00:11:14]** to get accepted then in the YouTube

**[00:11:16]** resources tab there'll be this video and

**[00:11:17]** we'll have all of the resources here

**[00:11:18]** including the the perfect prompt

**[00:11:20]** generator tool um access to the software

**[00:11:22]** that I'm going to be showing you in a

**[00:11:23]** second and also uh this document here

**[00:11:26]** and the prompt planning sheet Etc so

**[00:11:27]** that's all on the school community and

**[00:11:29]** here I've generated some example

**[00:11:30]** responses so as you can see uh we are

**[00:11:32]** expecting inputs of customer email and

**[00:11:35]** number of days since purchase so the

**[00:11:36]** inputs we have this QA format QA QA

**[00:11:39]** question answer question answer this

**[00:11:41]** will make a bit more sense when we do

**[00:11:42]** the drafting step next but basically we

**[00:11:44]** have the input of the email and we have

**[00:11:47]** um the number of days since purchase and

**[00:11:49]** this is the expected output so we have

**[00:11:51]** one pair here and an example input and

**[00:11:54]** output and then we have another pair

**[00:11:56]** here slightly different and the example

**[00:11:58]** output

**[00:11:59]** and the input here and another output

**[00:12:02]** example as well so these kinds of input

**[00:12:04]** and output pairs are really essential to

**[00:12:06]** make sure that the output is coming in

**[00:12:07]** the right format right tone of voice um

**[00:12:09]** and you guys will see in a second how

**[00:12:10]** when we bake this into the prompt we can

**[00:12:12]** even turn it on and off and see how much

**[00:12:13]** of a difference this makes so we have

**[00:12:15]** our examples here so we can go back to

**[00:12:17]** our prompting sheet um we can say client

**[00:12:19]** has provided they are included in the

**[00:12:23]** brief and can this be done in one step

**[00:12:25]** now in this case I'd say yes but

**[00:12:27]** sometimes you may need to chain multiple

**[00:12:28]** prompts together for example you may

**[00:12:30]** want to throw a whole bunch of

**[00:12:31]** information in and get it to write an

**[00:12:32]** outline and then it needs to go over

**[00:12:34]** that outline and kind of approve that it

**[00:12:35]** meets some kind of criteria or way that

**[00:12:37]** you like to structure your videos and

**[00:12:39]** then you need to do another step after

**[00:12:40]** that that's going to actually flesh it

**[00:12:41]** out into say a script or or into some

**[00:12:43]** piece of content that you want to write

**[00:12:44]** so sometimes it's called prompt chaining

**[00:12:46]** and you want to do multiple steps where

**[00:12:48]** it's you're asking a bit too much for it

**[00:12:49]** to do it all in one step but in this

**[00:12:51]** case we can do the classification of

**[00:12:52]** what type it is and also the generation

**[00:12:54]** of the response Allin one go you could

**[00:12:56]** split that out into okay let's do a

**[00:12:58]** classifier first that's more specialized

**[00:12:59]** in the writing of the responses rather

**[00:13:01]** than of the classification but in this

**[00:13:03]** case I think it's Overkill can this be

**[00:13:05]** done in one step yes and which model

**[00:13:07]** should be used this really comes down to

**[00:13:08]** a number of factors like how much is

**[00:13:10]** this going to be used if this is going

**[00:13:11]** to be hitting the API maybe thousands of

**[00:13:14]** times a month when it comes to the model

**[00:13:15]** we need to be thinking about what is

**[00:13:17]** this task for what context is it and how

**[00:13:19]** many times is it going to be used how

**[00:13:21]** difficult is the task so for more

**[00:13:23]** difficult tasks you're obviously going

**[00:13:24]** to want to go for the more expensive and

**[00:13:26]** and smarter models which typically do

**[00:13:28]** cost more but at the at the same time if

**[00:13:30]** this thing is going to be running

**[00:13:31]** thousands and thousands of times a month

**[00:13:33]** you may have a bit of a trade-off where

**[00:13:34]** hey we need a really intelligent model

**[00:13:36]** and we needed a lot so what's the best

**[00:13:38]** trade-off here how can I get it to work

**[00:13:39]** with the cheapest model possible or if

**[00:13:41]** it's only going to be used maybe a

**[00:13:42]** handful or a few dozen times a month

**[00:13:44]** then you may as well just go for a very

**[00:13:46]** smart model and just get the best

**[00:13:47]** results from it so in the cases where

**[00:13:48]** there's super high volume you're going

**[00:13:50]** to want to try and get the cheapest

**[00:13:51]** model possible but that can make a

**[00:13:52]** little bit more difficult and require a

**[00:13:54]** bit more prompt engineering but that's

**[00:13:55]** really where the value of of being a

**[00:13:58]** good prompt engineer and and having a

**[00:13:59]** good process comes from which is you can

**[00:14:00]** get better performance out of Cheaper

**[00:14:02]** models where other people have to rely

**[00:14:04]** on really really expensive ones you can

**[00:14:05]** use much cheaper ones with better

**[00:14:07]** instructions and get the same or even a

**[00:14:08]** better result so in this case because it

**[00:14:10]** is a kind of writing task I'm going to

**[00:14:11]** go with uh

**[00:14:13]** Claude son it 3.7 which is come out

**[00:14:16]** recently which is the best smartest

**[00:14:18]** writing model but that is really just

**[00:14:20]** the starting point we can Benchmark it

**[00:14:21]** with that and then we can try to go down

**[00:14:23]** to cheaper and cheaper models even

**[00:14:24]** within the CLA family but we want to

**[00:14:26]** Benchmark it and see just how good we

**[00:14:27]** can get it and then if we can use our

**[00:14:29]** engineering to work that model down to

**[00:14:30]** something cheaper and cheaper and does

**[00:14:31]** the client have specific requests none

**[00:14:34]** aside from in the brief okay so step two

**[00:14:37]** of the planning sheet is actually

**[00:14:38]** requirement Gathering so this is

**[00:14:39]** prepping everything that we're going to

**[00:14:40]** be putting into our AI tool in a second

**[00:14:42]** and a lot of this I'm just going to be a

**[00:14:44]** to grab directly from our sheet but this

**[00:14:45]** is really where you collect a lot of the

**[00:14:46]** information that's going to allow you to

**[00:14:48]** rapidly draft it in the next step so the

**[00:14:50]** purpose of this prompt I can go back to

**[00:14:52]** here maybe grab all of this and just

**[00:14:55]** take it back over inputs again this is

**[00:14:57]** similar to what we had just up here we

**[00:14:59]** can just grab

**[00:15:02]** these and a bit of context so this is

**[00:15:05]** the email and then days of purchase

**[00:15:07]** number of days since the cost be used

**[00:15:13]** for this is to be used for determining

**[00:15:14]** actions um such as refund request and

**[00:15:18]** now that we're at this IP section it's

**[00:15:20]** asking for the secret Source or

**[00:15:21]** methodology which we have in the brief

**[00:15:23]** here so if we go response categories and

**[00:15:26]** guidelines copy all of this

**[00:15:31]** we actually had this information here um

**[00:15:33]** maybe I'll just grab that and put it in

**[00:15:35]** as well this is the IP or secret

**[00:15:39]** source and for the output we can just do

**[00:15:41]** similar to what we had before so this

**[00:15:43]** step one of exploration is just to get

**[00:15:44]** your thoughts out on the page and make

**[00:15:45]** sure that you know what you're doing

**[00:15:46]** when you head into the second phase here

**[00:15:49]** so if we go down

**[00:15:52]** and plain text emo response and then for

**[00:15:54]** the examples we can go back down and

**[00:15:56]** grab our examples from the a client

**[00:16:05]** brief well we've got that there and at

**[00:16:07]** the bottom of the sheet there's actually

**[00:16:08]** another finished example so that if you

**[00:16:09]** are doing another prompt you can look at

**[00:16:11]** this and get an idea of how you can fill

**[00:16:12]** that out for a different kind of prompt

**[00:16:13]** and as I said this is from accelerator

**[00:16:15]** so there's some other stuff here and the

**[00:16:16]** next step is to use our plan in

**[00:16:18]** generating our draft which we're going

**[00:16:19]** to be covering in a second using our

**[00:16:21]** perfect prompt generator tools since

**[00:16:22]** we've filled out this entire requirement

**[00:16:23]** Gathering section we have completed our

**[00:16:25]** planning and now we're ready to move on

**[00:16:27]** to the drafting using our AI tool the

**[00:16:29]** perfect PR

**[00:16:31]** generator so to do the drafting step we

**[00:16:33]** need to sign up to relevance AI uh

**[00:16:35]** there'll be a link here in the resource

**[00:16:36]** so this will be on the school Community

**[00:16:38]** um and you can sign up to relevant Ai

**[00:16:40]** and once you have you can click this

**[00:16:41]** link which is going to allow you to

**[00:16:43]** clone a copy of our perfect prom

**[00:16:45]** generator into your account and up in

**[00:16:47]** the top right hand corner you'll see the

**[00:16:48]** Clone button so you want to click this

**[00:16:50]** and clone it into your account that

**[00:16:51]** means you are going to be responsible

**[00:16:52]** for uh paying the billing for it they've

**[00:16:54]** got a really good free plan so uh you

**[00:16:56]** don't need to worry about that and

**[00:16:57]** before we jump into using this and using

**[00:16:59]** the plan that we've just written to

**[00:17:00]** generate this prompt in a few seconds I

**[00:17:02]** just want to go behind the scenes a

**[00:17:03]** little bit and you guys will be able to

**[00:17:04]** check this out as well once it's in your

**[00:17:06]** account but you can see how I've put

**[00:17:07]** this tool together so this is what I

**[00:17:09]** want you guys to be able to do with the

**[00:17:10]** prompts that you create so whether

**[00:17:11]** you're an employee whether you're an

**[00:17:12]** agency owner whether you're a business

**[00:17:14]** owner using relevance AI to create your

**[00:17:15]** own tools like this this is an AI tool

**[00:17:18]** it's weird because we're using an AI

**[00:17:19]** prompt to create an AI tool that creates

**[00:17:21]** AI prompts can be a little bit tricky to

**[00:17:23]** wrap your head around but this is our

**[00:17:25]** perfect prompt generator tool and it's

**[00:17:26]** going to take in the use case

**[00:17:27]** information which is lot of what we just

**[00:17:29]** wrote in the planning and it's also

**[00:17:30]** going to take in some input and output

**[00:17:32]** examples when we click run on this it's

**[00:17:33]** going to send it to a an LM step here

**[00:17:36]** where I have baked in all of the latest

**[00:17:38]** research such as role prompting uh such

**[00:17:40]** as Chain of Thought prompting you have

**[00:17:42]** emotion prompting uh use of markdown

**[00:17:44]** formatting Etc uh fucha prompting Etc

**[00:17:47]** all of the information is already baked

**[00:17:48]** into this so what happens when you click

**[00:17:49]** on run this tool what it does is it

**[00:17:51]** passes in all of the information we did

**[00:17:53]** in the plan passes it into this and in a

**[00:17:55]** few seconds it applies all of those best

**[00:17:57]** practices for prompting to to the uh the

**[00:17:59]** inputs and to the use case that you've

**[00:18:01]** explained in the inputs so I might

**[00:18:03]** actually just change this to uh CL 3.7

**[00:18:06]** and so now you're ready to see just how

**[00:18:07]** quickly this uh prompt can be right in

**[00:18:09]** the draft so instead of having to do

**[00:18:10]** this all manually um you get to do it in

**[00:18:12]** a few seconds so if we come back to our

**[00:18:14]** planning sheet you go back up what you

**[00:18:16]** want to do is take um all of this first

**[00:18:19]** section from the purpose down

**[00:18:23]** to the examples so we just grab the

**[00:18:25]** output go back to relevance Ai and the

**[00:18:27]** perfect promp generator tool if we paste

**[00:18:29]** this in and that's all of this use case

**[00:18:31]** information what is the prompt doing in

**[00:18:32]** why give as much information as possible

**[00:18:35]** and then we can go down to our input and

**[00:18:36]** output examples and we can grab it here

**[00:18:39]** go

**[00:18:42]** down and if we paste this in here now

**[00:18:45]** and click run if we give it a few

**[00:18:47]** seconds it's going to write the full

**[00:18:49]** draft for this prompt and it's analyzing

**[00:18:51]** the use case all the information we

**[00:18:53]** provided on it all the rules that we

**[00:18:54]** gave it and that IP that we talked about

**[00:18:56]** and then it's applying all of the best

**[00:18:57]** prompting practices that are research

**[00:18:59]** back that we've included in the perfect

**[00:19:01]** prompt generator tool just here and now

**[00:19:03]** if we scroll down you can see that we've

**[00:19:04]** got the full prompt written out so let's

**[00:19:06]** just take a scem through it right we

**[00:19:07]** have a role you are an expert customer

**[00:19:08]** support autor responded for Te Gear Pro

**[00:19:11]** R skilled in analyzing customer

**[00:19:12]** inquiries identifying to issue types and

**[00:19:14]** crafting personalized help with

**[00:19:16]** responses that either resolve issues

**[00:19:17]** completely or set appropriate

**[00:19:19]** expectations for human followup

**[00:19:20]** responses are friendly concise R task

**[00:19:22]** and the task it seems to have nailed

**[00:19:24]** which is analyze the incoming customer

**[00:19:25]** email and provide an appropriate

**[00:19:27]** response using the step-by-step process

**[00:19:28]** so this is Chain of Thought prompting

**[00:19:30]** which it's applied carefully read the

**[00:19:31]** customer's email identify the order type

**[00:19:34]** consider the days since the purchase

**[00:19:35]** determine if the issue can be completely

**[00:19:37]** resolved Etc so it's taken all of that

**[00:19:39]** IP and that juice and instructions that

**[00:19:41]** we've provided and it's applied it

**[00:19:43]** through a a Chain of Thought method in

**[00:19:44]** this case and then here we have things

**[00:19:46]** like emotion prompting which is saying

**[00:19:47]** we greatly value your ability to strike

**[00:19:49]** the perfect balance your expertise is is

**[00:19:51]** really really appreciated and this is

**[00:19:52]** proven uh through research to improve

**[00:19:55]** the performance of the prompts versus

**[00:19:56]** when you don't have these emotional

**[00:19:57]** words in it as well so it's all woven in

**[00:20:00]** there all of the best and latest

**[00:20:01]** research and prompt engineering and here

**[00:20:03]** we have the inputs that it's going to be

**[00:20:05]** expecting and then also as context it's

**[00:20:06]** given all of the rules um that the AI

**[00:20:08]** needs to be following when responding

**[00:20:11]** and it's got our examples in here too um

**[00:20:13]** the layout seems to be a little bit off

**[00:20:15]** but um it's also got the notes section

**[00:20:17]** down the end and what we want to do from

**[00:20:19]** here we have technically done the

**[00:20:20]** drafting as you can see that's really

**[00:20:22]** where we put a lot of effort into the

**[00:20:23]** plan because once you throw it into here

**[00:20:25]** um and we click this format it then we

**[00:20:27]** have it all ready to go nicely formatted

**[00:20:29]** and markdown formatting which is another

**[00:20:31]** thing that helps with prompt performance

**[00:20:33]** and we have it all nicely structured and

**[00:20:35]** ready to go into the next step so that

**[00:20:36]** is drafting and now we can jump into

**[00:20:38]** actually evaluating using our prompt

**[00:20:40]** engineering software called prompt

**[00:20:43]** metheus okay so here we are this is the

**[00:20:45]** software that we use at Morningside for

**[00:20:47]** prompt engineering and this has been in

**[00:20:49]** sort of secret weapon of mine and the

**[00:20:50]** team for a long time I have shared it

**[00:20:52]** with my accelerator members but now I

**[00:20:54]** think it's time to to probably share uh

**[00:20:55]** with you all on the channel here it's

**[00:20:57]** what's called a prompt engineering ID e

**[00:20:59]** which is an integrated development

**[00:21:01]** environment and this is what um coders

**[00:21:03]** or developers use when they are writing

**[00:21:04]** software so you may be familiar with vs

**[00:21:07]** code I'll put some stuff up on screen it

**[00:21:08]** may ring a few BS if you've ever watched

**[00:21:10]** any kind of tutorials but applications

**[00:21:12]** like vs code are basically a a

**[00:21:13]** all-in-one development software that

**[00:21:16]** allows developers to easily write code

**[00:21:18]** manage files uh run code via the

**[00:21:20]** terminal it's all all the tools you need

**[00:21:22]** really in one place it's the integrated

**[00:21:24]** development environment and this is a

**[00:21:25]** prompt engineering integrated

**[00:21:27]** development environment with all of the

**[00:21:29]** best tools that you need to be able to

**[00:21:31]** write and test and improve prompts like

**[00:21:34]** a like an engineer and taking more of a

**[00:21:36]** scientific approach rather than sort of

**[00:21:37]** an artsy approach which is what many

**[00:21:39]** people do to some success of course I

**[00:21:41]** know this is not a sponsored video I'm

**[00:21:42]** not getting paid by anyone to promote

**[00:21:43]** this or relevance um this is just

**[00:21:45]** showing you exactly what we do at

**[00:21:47]** morning side so uh the link to sign up

**[00:21:48]** to Prometheus will be with all the other

**[00:21:50]** links and resources on the school so you

**[00:21:52]** can get it there and sign up and once

**[00:21:53]** you're inside you'll see something that

**[00:21:54]** looks a bit like this and so on the left

**[00:21:56]** side here we have our prompts and data

**[00:21:58]** sets I'm just going to click add a new

**[00:21:59]** prompt here and we can start to build

**[00:22:01]** out our prompt within prompt metheus so

**[00:22:03]** that we can start to test it and uh

**[00:22:05]** evaluate we are in the evaluate part of

**[00:22:07]** the Peter process now which means we

**[00:22:09]** need to load up our prompt into prompt

**[00:22:10]** metheus we need to run some inputs

**[00:22:12]** against it and evaluate where we are

**[00:22:14]** like how far are we off where we're

**[00:22:16]** actually trying to get to is the tone

**[00:22:18]** where we need it to be is the prompt

**[00:22:19]** actually getting the classification

**[00:22:20]** correct so testing all of these things

**[00:22:22]** is what the evaluation step is with this

**[00:22:24]** initial draft so to get things set up in

**[00:22:26]** PR metheus it's actually really really

**[00:22:27]** easy if you just go back to relevance Ai

**[00:22:29]** and you copy these sections if I copy

**[00:22:32]** the roll here and I go back and then I'm

**[00:22:34]** going to add this block as

**[00:22:37]** roll and you'll see why we start to

**[00:22:39]** separate them into these different

**[00:22:40]** chunks as it makes it a lot easier for

**[00:22:41]** us to test things so we have our roll in

**[00:22:44]** here I'm going to add another block

**[00:22:46]** here that's the task this

**[00:22:49]** time I am going to leave that off

**[00:23:03]** I'm just going to skip over the data

**[00:23:04]** here for now because we need to do

**[00:23:05]** something that's a little bit different

**[00:23:06]** for this um which is going to take in

**[00:23:08]** the various different inputs we want to

**[00:23:09]** test so we can get specifics

**[00:23:23]** now now grabbing the context

**[00:23:32]** so what you can see me doing here is

**[00:23:33]** putting the markdown formatting in so

**[00:23:35]** this just helps the AI to understand the

**[00:23:36]** structure of the prompt better so this

**[00:23:38]** is it's reading this as a heading and

**[00:23:39]** this is a like an H2 or or a heading two

**[00:23:43]** um so it gives it a lot more structure

**[00:23:45]** so this is a specific section this is

**[00:23:46]** the task this is the role Etc and here

**[00:23:49]** we have a uh a sort of Hing to under the

**[00:23:51]** context which is all about the uh the

**[00:23:52]** response categories and guidelines as

**[00:23:54]** well so now I can do the same with the

**[00:23:55]** examples

**[00:24:18]** and now we could just grab the notes

**[00:24:19]** Here

**[00:24:31]** we have all the different components of

**[00:24:32]** the prompt split up into the different

**[00:24:33]** sections so that's going to make it very

**[00:24:34]** easy for us to test in a second but the

**[00:24:36]** last step before we actually start

**[00:24:37]** evaluating is we need to go and insert

**[00:24:40]** the variable or the data set like I said

**[00:24:41]** earlier we want to be able to test this

**[00:24:43]** against multiple different inputs and

**[00:24:45]** determine if it's working as expected so

**[00:24:46]** promp neus allows us to do this very

**[00:24:48]** easily by going into what's called data

**[00:24:50]** sets and if I go back to the uh

**[00:24:53]** information that was provided by the

**[00:24:54]** client or the brief for the project I

**[00:24:56]** have all of the input data here so we

**[00:24:58]** have these

**[00:24:59]** peers we can go this is the email and

**[00:25:02]** then the day since purchases in there as

**[00:25:04]** well so if you look at this aside from

**[00:25:05]** the from in the subject which you could

**[00:25:07]** probably kill and and take out we have

**[00:25:09]** the same format in the examples here so

**[00:25:10]** the email and then the day since

**[00:25:12]** purchase so that's the input format that

**[00:25:14]** is expecting and also just to make it a

**[00:25:16]** bit easier we can say rename this here

**[00:25:18]** with this little rename button on the

**[00:25:19]** side you can say this is our

**[00:25:22]** um Tech Gear

**[00:25:25]** Pro email we can also rename our data

**[00:25:28]** dat set here um Tech Gear

**[00:25:32]** Pro and so this is one item in the data

**[00:25:34]** set this will all make sense in a second

**[00:25:35]** so stick with me if we then go add

**[00:25:37]** another

**[00:25:38]** item we go back to

**[00:25:43]** this add another item so I've got 10 of

**[00:25:45]** these to add in um these are going to be

**[00:25:46]** on the resource you can copy them in if

**[00:25:48]** you're following along but the idea is

**[00:25:50]** the same that uh you want to have a

**[00:25:51]** bunch of different cases that we can

**[00:25:52]** test against and I'm just going to sort

**[00:25:55]** of go off camera and just add these in

**[00:25:56]** now all right so now I have 10 items in

**[00:25:58]** my data set and these are dummy emails

**[00:26:00]** or dummy data sets that are going to be

**[00:26:02]** passed to The Prompt um or hypothetical

**[00:26:04]** ones that it should be able to deal with

**[00:26:05]** and sort of trying to simulate as much

**[00:26:07]** as possible what it would look like in

**[00:26:08]** the real world so now we can go back to

**[00:26:10]** our uh te Gear Pro email autoresponder

**[00:26:12]** prompt and we need to insert into here a

**[00:26:16]** data block so we can click on Tech

**[00:26:17]** gearpro here and as you can see we can

**[00:26:20]** now cycle through all these different

**[00:26:22]** cases to test which is super great um

**[00:26:25]** and here what we want to do is add on

**[00:26:26]** the end if we go back to relevance AI

**[00:26:28]** can see that we did have a little bit

**[00:26:31]** here and this is basically where it was

**[00:26:32]** saying we should input the variables

**[00:26:34]** what I like to do is add a little bit

**[00:26:35]** more uh we can go like um

**[00:26:40]** input data and so this stuff here is

**[00:26:42]** obviously going to slot right underneath

**[00:26:43]** whatever we have here so we can say

**[00:26:45]** classify and

**[00:26:50]** write and then we have it now when we

**[00:26:53]** run this it's going to fit just under

**[00:26:55]** here and this is going to be the uh

**[00:26:57]** input data section so I'll click back to

**[00:26:58]** one here and we almost almost ready all

**[00:27:01]** we need to do now is to select the model

**[00:27:02]** so as I said I wanted to run a a clae

**[00:27:05]** 3.7 which is the latest one they have

**[00:27:07]** gives us the prices here which is super

**[00:27:09]** helpful when you are planning these out

**[00:27:10]** it has all of the options here for

**[00:27:12]** temperature um I might just put this

**[00:27:13]** down to 0.1 for temperature and this is

**[00:27:15]** something we're going to be testing in

**[00:27:16]** the next step and you have things like

**[00:27:18]** the token limit um I would probably chop

**[00:27:20]** that down to 400 there's things like

**[00:27:21]** Json mode so remember when we talked

**[00:27:23]** about how you can have different kinds

**[00:27:24]** of outputs Json mode allows you to play

**[00:27:25]** around with that and and sort of strict

**[00:27:27]** and force a Jon formatted output and

**[00:27:29]** before we actually run this you will

**[00:27:30]** need to set up your API keys so you can

**[00:27:32]** come to your account down here by

**[00:27:34]** clicking on the bottom left and manage

**[00:27:36]** API Keys you can see I've got my

**[00:27:38]** anthropic one in there you can go to

**[00:27:39]** console. anthropic tocom and you can set

**[00:27:41]** up your account um and get your API key

**[00:27:43]** and pop it in here you can do it for all

**[00:27:44]** these different providers deep seek

**[00:27:46]** Gemini grock mistol open AI I've got

**[00:27:48]** mine in there too xai and everything so

**[00:27:51]** in order to use the models you obviously

**[00:27:52]** need to access it through their API but

**[00:27:54]** this is how you can manage your API Keys

**[00:27:55]** within prom metheus so now we are ready

**[00:27:58]** to give this a spin we have all the data

**[00:28:00]** in we are going to be testing this first

**[00:28:02]** case which is I ordered the premium

**[00:28:03]** wireless earbuds on Monday and was

**[00:28:05]** wondering when they were arrive I paid

**[00:28:07]** for standard shipping and I'm just

**[00:28:08]** curious about the timeline so this is

**[00:28:09]** someone who purchased recently and is

**[00:28:10]** wondering where their product is it

**[00:28:12]** should be able to correctly classify

**[00:28:13]** this as an email about an order status

**[00:28:15]** and then realize that since it's just 4

**[00:28:17]** days since purchase that they should

**[00:28:18]** just be sort of prompted to wait a

**[00:28:19]** little bit longer so if we give this a

**[00:28:21]** spin all right there we go we have our

**[00:28:22]** first response and it's saying after

**[00:28:24]** analyzing the customer email I can see

**[00:28:25]** this so first thing we've noticed

**[00:28:27]** already as we test the first one is that

**[00:28:28]** it's doing this typical AI thing with

**[00:28:30]** the of conversational chat models we

**[00:28:32]** have these days it's it's writing a bit

**[00:28:34]** of a lead into it so right there that's

**[00:28:35]** something that I'd want to tweak and

**[00:28:36]** that's going to be in the next step of

**[00:28:38]** refining at this point you'd want to be

**[00:28:39]** making some notes of the different

**[00:28:41]** issues that you want to solve and say

**[00:28:42]** hey look I want to add down in the note

**[00:28:44]** section what I want to do is add in a a

**[00:28:46]** reminder that they should only be

**[00:28:47]** replying with the email there should be

**[00:28:49]** no other context provided just the email

**[00:28:51]** is the output but it also does depend on

**[00:28:53]** the system if this was not fully

**[00:28:54]** automated and then the AI wasn't

**[00:28:56]** actually going to send this email

**[00:28:57]** automatically

**[00:28:58]** then having this a bit of a summary and

**[00:29:00]** sort of Chain of Thought of it

**[00:29:01]** explaining why it's doing or why it

**[00:29:03]** wrote what it wrote can actually be

**[00:29:05]** quite helpful because it increases the

**[00:29:06]** accuracy of it but in this case we're

**[00:29:07]** assuming that this is going to be for an

**[00:29:09]** automated response system and so you

**[00:29:11]** wouldn't want it to attack any of that

**[00:29:13]** on so it's very important that we prompt

**[00:29:14]** that out but for now we can see that

**[00:29:16]** it's getting uh the classification and

**[00:29:18]** the response right based off our

**[00:29:20]** instructions what you can do in PR

**[00:29:21]** metheus is grade them very easily

**[00:29:23]** because there's the stuff at the top

**[00:29:24]** I've given it a thumbs down and this is

**[00:29:25]** going to allow me to see over time which

**[00:29:27]** ones are good responses of which are

**[00:29:29]** this one is a missing order but it's

**[00:29:30]** taken longer than 10 days so we should

**[00:29:33]** be able to see it's saying hey look let

**[00:29:34]** me go and get a human on

**[00:29:37]** this there we go it's saying um this

**[00:29:40]** delay is unusual I'm personally looking

**[00:29:42]** into what happens R I'll keep you

**[00:29:44]** updated over the next 24 so it's

**[00:29:45]** basically buying time for the human team

**[00:29:48]** and again this is probably a thumbs down

**[00:29:50]** because it's got the stuff at the start

**[00:29:51]** and I think already we can start to see

**[00:29:53]** where we want to be tweaking The Prompt

**[00:29:55]** and the direction that we want to be

**[00:29:56]** making changes so we can jump into the

**[00:29:57]** next tip which is going to be the refine

**[00:30:01]** stage so in this refining process this

**[00:30:03]** is a an iterative thing that you do sort

**[00:30:04]** of over and over it goes evaluate and

**[00:30:06]** refine and evaluate where you're making

**[00:30:07]** changes to the prompt evaluating them

**[00:30:09]** again and it's a cycle that you go

**[00:30:10]** through until you're happy with the

**[00:30:11]** results so in this case as I said we're

**[00:30:13]** getting this filler or this uh this

**[00:30:15]** Preamble um at the start of all of them

**[00:30:17]** which we want to immediately try to buff

**[00:30:19]** out um which we can go likely to say the

**[00:30:22]** task section here um actually let's go

**[00:30:25]** to the specific section I'm going to

**[00:30:27]** create a new version so this is one of

**[00:30:28]** the tools that we have within prop

**[00:30:30]** metheus is that we can add a variant so

**[00:30:32]** I'm going to create a variant of these

**[00:30:33]** specifics I'm actually going to go back

**[00:30:35]** and copy what we had I'm going to put it

**[00:30:38]** in here and say add one more to the

**[00:30:40]** bottom of this and say

**[00:30:43]** your response so I've said your response

**[00:30:45]** should not include anything other than

**[00:30:46]** the email to be sent to the customer do

**[00:30:48]** not include any other prbl reasoning

**[00:30:50]** this is

**[00:30:52]** because you are part of a an automated

**[00:30:57]** resp response system so Le you must not

**[00:31:03]** have this

**[00:31:04]** extra con content at the start of

**[00:31:09]** your responses and then I'll probably

**[00:31:12]** try to reinforce this again I might just

**[00:31:13]** copy it for now um I will make a

**[00:31:16]** duplicate of the context because I think

**[00:31:18]** in the uh in the context

**[00:31:20]** here we can make it a lot clearer that

**[00:31:22]** this is an automated system rather than

**[00:31:24]** one that it's allowed to do this kind of

**[00:31:25]** stuff at the start so there I've just

**[00:31:27]** added in a little bit of extra giving it

**[00:31:28]** even an example of the kind of stuff I'm

**[00:31:30]** talking about so I've taken a little

**[00:31:31]** excerpt out of this and say adding

**[00:31:33]** reasoning or any other waffle like XYZ

**[00:31:35]** could be disastrous because it would

**[00:31:37]** reveal that the system is automated so I

**[00:31:39]** have a new version of my context here um

**[00:31:41]** I could tack it on to the end of the

**[00:31:42]** notes but we have our new version of the

**[00:31:44]** specifics with this one added at the end

**[00:31:45]** and a new version of the context so now

**[00:31:48]** we can give this a run I'll probably run

**[00:31:49]** it back over the first two that we just

**[00:31:51]** did to make sure that we''ve got them

**[00:31:53]** picked off there we go so we're getting

**[00:31:55]** the kind of responses we want I think

**[00:31:57]** the formatting could be improved but in

**[00:31:59]** the real world you'd be building this

**[00:32:00]** into some kind of workflow Automation

**[00:32:02]** and you could do an additional step

**[00:32:03]** after that's a formatter that preps it

**[00:32:04]** for say the Gmail or whatever email

**[00:32:07]** system that you're using to send and it

**[00:32:08]** can make sure that it's actually

**[00:32:09]** formatted correctly but in this case uh

**[00:32:11]** we've just got the the content correct

**[00:32:13]** and the classification correct so now we

**[00:32:15]** can give this a nice thumbs up and we

**[00:32:17]** can go to the next one and run it again

**[00:32:19]** and the goal here is to be able to go

**[00:32:21]** through all of these um examples the

**[00:32:23]** input examples that I've put into the

**[00:32:24]** data set and for it to be categorizing

**[00:32:26]** correctly and giving the right kinds of

**[00:32:28]** responses based on the examples and

**[00:32:30]** based on the uh the guidelines that the

**[00:32:32]** customer or the client is given us yep

**[00:32:33]** so this one's saying that they're going

**[00:32:34]** to be looking into it with than 24 hours

**[00:32:37]** that's all good and we can go through to

**[00:32:40]** number three

**[00:32:41]** here so I think it's pretty clear what

**[00:32:43]** the refinement process is from here and

**[00:32:45]** the evaluation refinement Loop um you go

**[00:32:47]** back and forth until you get in the kind

**[00:32:48]** of results that you're looking for I

**[00:32:49]** want to give you a few more tips here um

**[00:32:52]** just so you can play around more levers

**[00:32:53]** that you can pull and there are a few

**[00:32:55]** big ones here and this actually really

**[00:32:56]** helps you to understand prompt

**[00:32:57]** engineering on on a much deeper level

**[00:32:59]** because say here we have the switch

**[00:33:00]** where we can toggle things on and off so

**[00:33:02]** entire sections can be toggled on and

**[00:33:03]** off so we can do some interesting

**[00:33:05]** experiments like okay if I turn off the

**[00:33:07]** roll section and then I run this again

**[00:33:10]** is it going to come out much different

**[00:33:11]** to what it just

**[00:33:14]** did so looks like it's pretty much the

**[00:33:18]** same there so it appears that the roll

**[00:33:19]** is not actually adding much to it in

**[00:33:21]** this case likely because we've got so

**[00:33:22]** much context elsewhere but this is a

**[00:33:24]** fairly straightforward use case and we

**[00:33:25]** are using quite a powerful model which

**[00:33:28]** is something we can test in a second

**[00:33:29]** actually is bringing that model quality

**[00:33:30]** down and seeing if we can get the same

**[00:33:32]** results with an even cheaper model um or

**[00:33:34]** another good one that I like to do is

**[00:33:35]** turning off the examples and seeing if

**[00:33:37]** they're actually making a difference

**[00:33:38]** because in a lot of cases your examples

**[00:33:40]** are going to be a massive massive amount

**[00:33:42]** wall of text um especially when you have

**[00:33:44]** large inputs um and you maybe have a a

**[00:33:47]** whole V YouTube video that you want to

**[00:33:48]** put in and the transcript is the input

**[00:33:50]** and and therefore your example section

**[00:33:51]** is gigantic um but in this case let's

**[00:33:54]** just turn off our examples and we'll

**[00:33:55]** give it a spin and see how this one

**[00:33:57]** comes out I actually there is a bit of a

**[00:33:59]** difference see here we have the roll and

**[00:34:00]** the task on and it gave us this nice

**[00:34:02]** nicely formatted exactly how we wanted

**[00:34:04]** it and then if we go up we have taken

**[00:34:06]** the roll out and one we getting the

**[00:34:08]** subject in there so it seems to be a lot

**[00:34:10]** less accurate it's not following the

**[00:34:11]** exact response format that we've asked

**[00:34:12]** for here which is it's obviously

**[00:34:13]** ignoring the fact that we said look you

**[00:34:15]** you can't put in any other stuff that's

**[00:34:17]** going to reveal that this is not a

**[00:34:18]** respond so this would be a fail this one

**[00:34:20]** would be good and here you can see we've

**[00:34:23]** added the roll back in but taken out the

**[00:34:25]** examples and we're seeing similar issues

**[00:34:27]** again so the role in this case actually

**[00:34:28]** appears to be doing quite a lot it's

**[00:34:30]** paying a lot more attention to the

**[00:34:31]** instructions we give it and also

**[00:34:32]** understanding that it's part of an

**[00:34:33]** autoresponder system and therefore this

**[00:34:35]** stuff can't be included it appears that

**[00:34:37]** the example section where we have it

**[00:34:38]** here but not in here is not actually

**[00:34:40]** doing that much so you can see how you

**[00:34:42]** can start to test and figure out what

**[00:34:44]** parts of the prompt are actually moving

**[00:34:45]** the needle what parts are needed um that

**[00:34:47]** actually isn't isn't a green thumb so if

**[00:34:49]** we turn everything back

**[00:34:50]** on I've shown you guys how you can

**[00:34:52]** create new versions of sections easiest

**[00:34:54]** way is to right click on one of these

**[00:34:56]** tabs and click duplicate and it's going

**[00:34:57]** to give you a and a duplicate version of

**[00:34:59]** that you can make changes to it chops

**[00:35:00]** and bits out and the last test that I'd

**[00:35:02]** want to do here is see if we can get the

**[00:35:03]** same performance um let's just run

**[00:35:05]** through a couple of these there we go um

**[00:35:11]** four five and six so we'll just get

**[00:35:16]** these filled out using the uh the best

**[00:35:18]** prompt that we know is is performing how

**[00:35:20]** we want it to and there we go we have

**[00:35:21]** all of these responses let's just double

**[00:35:23]** check this last one number six this is a

**[00:35:25]** technical support one and it seems to be

**[00:35:26]** dealing with that correctly there so

**[00:35:27]** that's good this is performing as

**[00:35:29]** expected now if we change this down to a

**[00:35:32]** cheaper version of claude's model so 3.5

**[00:35:35]** hiu and we run this on six here and then

**[00:35:38]** we compare

**[00:35:39]** them now actually this probably isn't a

**[00:35:41]** fair comparison because uh this part

**[00:35:43]** with this tech support kind of questions

**[00:35:45]** it is coming up with its own AI

**[00:35:46]** generated responses to we don't actually

**[00:35:48]** have a knowledge base on tech support

**[00:35:49]** which is what you do with more of an

**[00:35:50]** agent based build but if we go back to

**[00:35:53]** um let's just do one two and three again

**[00:35:55]** so an easy way of doing this would be to

**[00:35:56]** just go to High and we'll run

**[00:35:59]** one and then we'll switch to Sonet 3.7

**[00:36:02]** and we'll compare it side by

**[00:36:05]** side and it appears that we are

**[00:36:08]** getting yeah pretty much the same thing

**[00:36:10]** and you can see the cost has gone from

**[00:36:11]** 83 cents here down to 24 cents it's

**[00:36:15]** taken almost a second less to get this

**[00:36:16]** same kind of result so prop metheus is

**[00:36:17]** great for giving you some helpful

**[00:36:18]** outputs like this obviously this is uh 3

**[00:36:21]** times cheaper um than this one and if we

**[00:36:24]** go for back to uh 3.5 here and we test

**[00:36:27]** number two and then we go to uh 3.7 we

**[00:36:31]** test number two again you can see okay

**[00:36:33]** we are

**[00:36:34]** getting yep pretty much the same

**[00:36:36]** response um exactly the same on both of

**[00:36:38]** them there's also other things you can

**[00:36:39]** play with if you're doing more writing

**[00:36:40]** based tasks like come into the

**[00:36:42]** temperature and play around with the

**[00:36:43]** temperature but I think I've given you a

**[00:36:45]** pretty good rundown of the entire

**[00:36:46]** process of uh how we build prompts it's

**[00:36:48]** Peter process we write the plan out use

**[00:36:50]** that plan put it into the tool to

**[00:36:51]** rapidly generate the draft fit all the

**[00:36:53]** pieces into promt metheus here set up

**[00:36:56]** your data sets and then rapidly go

**[00:36:57]** through and test all of them to make

**[00:36:58]** sure it's working and that refinement

**[00:37:00]** cycle and the different levers you have

**[00:37:01]** to pull on like turning sections off

**[00:37:03]** creating new uh variations of certain

**[00:37:05]** sections playing around with the model

**[00:37:07]** the temperature etc those are all your

**[00:37:09]** tools and I want to just give you a

**[00:37:11]** final sort of Keystone or cap to this

**[00:37:13]** entire project which is how you can turn

**[00:37:15]** these prompts into uh usable tools that

**[00:37:17]** can actually be really helpful for you

**[00:37:19]** either as an employee or or or business

**[00:37:21]** owner is to actually go back to relevant

**[00:37:23]** Ai and to make handy internal tools for

**[00:37:26]** yourself or for other people in the

**[00:37:27]** company or even for clients if you're an

**[00:37:28]** AI Agency on relevance here say if you

**[00:37:31]** wanted to turn this into a autoresponder

**[00:37:33]** tool not a super great example but you'd

**[00:37:35]** probably do this more for say like uh

**[00:37:36]** content generation where you put in a

**[00:37:37]** couple of inputs and it gives you uh

**[00:37:39]** through your prompt engineering work

**[00:37:41]** it's going to be a to transform that

**[00:37:42]** into an ideal kind of content for

**[00:37:44]** whatever platform you're choosing but in

**[00:37:45]** this case we can then take all of these

**[00:37:47]** um this is of course taking in two

**[00:37:49]** inputs so we could say input one is

**[00:37:50]** going to be email

**[00:37:55]** content we can go a have be short text

**[00:37:58]** and we can go um days since per then we

**[00:38:03]** can go to an llm step here and what we

**[00:38:05]** can do is just start to build out our

**[00:38:07]** prompt again inside relevant Ai and then

**[00:38:10]** we can just pop these in here and go

**[00:38:11]** email

**[00:38:14]** content and put in the email there and

**[00:38:17]** if we go at days since

**[00:38:21]** purchase don't need to put this in of

**[00:38:23]** course cuz we're going to be passing

**[00:38:24]** that in within relevance and we can just

**[00:38:26]** dump the rest of the prompt there here

**[00:38:28]** now that we've finalized it and really

**[00:38:30]** worked it through so the thing is that

**[00:38:31]** testing it within something like

**[00:38:32]** relevance or any other kind of workflow

**[00:38:34]** Builder would be a lot more difficult

**[00:38:37]** and take a whole lot more time and you

**[00:38:39]** don't have anywhere near as much control

**[00:38:41]** um and it's just really an inferior way

**[00:38:43]** of testing and building the prompt so we

**[00:38:45]** can then replicate The Prompt that we

**[00:38:46]** chose here which is going to be our claw

**[00:38:48]** 3.5 hiu and if we save and

**[00:38:52]** publish we now have our own email reply

**[00:38:54]** generator tool so say if you want to

**[00:38:56]** make a tool like this use your your pro

**[00:38:57]** engineering skills or get your your

**[00:38:59]** staff or team to do this and then you

**[00:39:00]** can go to

**[00:39:02]** use how you could even go share this is

**[00:39:04]** usually how I'd like to share them

**[00:39:05]** around the

**[00:39:06]** company copy this and now you've got

**[00:39:09]** this cool Standalone tool where anyone

**[00:39:10]** in the company can access this tool and

**[00:39:13]** I can go back and do the exact same

**[00:39:15]** thing we did and we put in

**[00:39:18]** our where's my

**[00:39:22]** order and we've got four days since

**[00:39:25]** purchase

**[00:39:27]** and Bam that's how you can create your

**[00:39:29]** own handy dandy AI tools um and use them

**[00:39:32]** sort of wherever you need within your

**[00:39:33]** business within your personal life um

**[00:39:35]** and using that prompt Engineering

**[00:39:36]** Process and there you go we're getting

**[00:39:37]** the kind of outputs we want and this can

**[00:39:38]** be used by API all sorts of different

**[00:39:40]** things but that's the prompt Engineering

**[00:39:41]** Process first and foremost that I wanted

**[00:39:43]** to teach you and while that planning

**[00:39:44]** phase can feel a bit clunky at first

**[00:39:46]** it's so good if you just rip through

**[00:39:47]** that and then you paste that into the

**[00:39:48]** the AI tool to generate it and before

**[00:39:50]** you know it you've got a like 90% of the

**[00:39:52]** way there with a few tweaks needed as

**[00:39:54]** you saw but being able to extract more

**[00:39:55]** value out of these AI models and not

**[00:39:57]** having pay for that tradeoff of okay I

**[00:39:58]** don't have the time to spend to make a

**[00:40:00]** really good prompt this is the the whole

**[00:40:02]** system that we use at warning sides so I

**[00:40:03]** hope that's been helpful for you guys if

**[00:40:04]** you want to get all the resources and

**[00:40:05]** the perfect prompt generator tool I will

**[00:40:07]** even give you access to our perfect

**[00:40:08]** prompt generator tool for AI agents

**[00:40:10]** which is a bit more complex but I know a

**[00:40:12]** lot of you guys are building agents now

**[00:40:13]** and we have a really really good tool

**[00:40:14]** for doing that as well so that's all

**[00:40:15]** going to be linked in the first link in

**[00:40:17]** the description will be to our school

**[00:40:19]** you go to YouTube resources on school

**[00:40:20]** once you're inside and you'll find this

**[00:40:22]** video and all the resources will be on

**[00:40:23]** there also my entire free course as well

**[00:40:25]** for building AI businesses building AI

**[00:40:28]** systems building agents and everything

**[00:40:29]** is in there as well so that's all for

**[00:40:31]** the video guys if you want to check out

**[00:40:32]** my recent video which is massive

**[00:40:33]** breakdown of how to build AI agents like

**[00:40:35]** probably the best resource on the entire

**[00:40:37]** internet right now about how to learn AI

**[00:40:39]** agent building um if you're a

**[00:40:40]** non-technical person or even if you are

**[00:40:42]** a technical person that's up there but

**[00:40:43]** that's all for the video guys thank you

**[00:40:44]** so much for watching and I will see you

**[00:40:46]** in the next one

# Full Text (without timestamps)

over the past two years my AI agency morning side AI has built AI systems for companies of all sizes from local martial arts gyms to publicly traded companies and even an NBA team and at the core of every one of our client projects are a handful of powerful AI prompts which in a few hundred carefully chosen words can automate hundreds or even thousands of hours of manual work so being able to write really really good prompts like these is a superpower these days regardless of what career you're in but the tricky thing is that writing these kinds of really good prompts usually takes a lot of time and effort realizing this at Morningside AI it became a major priority for us to figure out how to write these kind of elite production grade prompts for our client projects as quickly as possible and so over the past 2 years using the latest in AI prompting research and professional grade prompting software we created our own rapid prompt engineering system called the Peter process and in this video I'm going to be revealing our entire system including the exact prompt engineering tools and software we use and the process as well that we use at Morningside AI to craft hyper effective prompts in just minutes that are able to extract the most value from AI models possible without having to spend all day writing and testing those prompts so we're going to start quickly with an overview of the Peter framework and then diving into a more Hands-On walkthrough of exactly what it looks like when it's in action and we're basically going to be building a business grade prompt from scratch using this process so whether you're an employee who's just keen on automating some of the work you do or you're an AI agency owner wanting to deliver better client results and more consistently or if you're a business owner and you're looking to train up your own team on how to write prompts well and write them fast and be able to automate their own work then this video will show you how the professionals do it so first and foremost let's just talk about what makes prompt engineering so powerful right it's kind of overused and over talked about but when done right it is essentially like having a magic wand that can transform data from one form into another you're taking some kind of raw information and then turning it into something valuable just by writing some magic words in the middle it's pretty magical when you think about it but here's the reality that most people face right there's a massive difference between casual prompting that most people do and professional prompt engineering which is what I'm going to teach you how to do in this video so when you are just chatting to chat gbt you can go back and forth and you can kind of refine your requests by asking follow-up prompts and sending more more questions than nudg in the right directions until you get what you want so it may take a few tries but eventually you will get to the outcome you want but in the real world whether you're Building Systems for businesses like myself and the team at morning side AI or you're creating some client deliverables or you're automating entire workflows you need to be able to prompt at a level of what's called single shot prompting basically your prompt needs to be able to work perfectly the first time every time without any human intervention and kind of going back and forth and saying hey look could you could you do it without these words or in this kind of tone right you don't have that back and forth it needs to be able to be given data transform that how you want it and do that reliably at scale and this is where all of the value really is for prompt engineering as a skill but this is also where most people get stuck in this frustrating trade-off that I mentioned earlier on one hand you can either write prompts quickly that perform okay or you can spend hours crafting the ones that work flawlessly and when you're trying to build something valuable whether that's for your employer or for your clients or for your own business neither option is acceptable let me give you some quick examples of why this matters so say you're an employee right imagine you are automating a weekly report that normally takes you 3 hours so a mediocre prompt that you do quickly might get you maybe 70% of the way there there's still an hour or so of cleanup work that you have to do whereas a great prompt that you spend a lot of time on might do 95% of the work for you and pretty much save you that entire 3 hours and for AI automation agency owners when you are delivering to your clients the difference between a good and a great prompt can mean the difference between a one-time project and a long-term retainer and relationship with thousands and thousands of dollars per month or if you're a business owner training your team to write Elite prompts means that you can Implement AI across your entire organization with consistent and reliable results so what was in realizing the difference between these good and great prompts and how much extra leg work you can get the AI systems to do with the great prompts but also factoring in that these take a lot longer to write that we decided to create our own process at morning side AI to always get the best prompts that get the most out of AI models without having to spend all day writing and testing them so the Peter process is a systematic approach that allows us to create these high performing prompts in a fraction of the time that it would normally take us it's so simple that anyone can be trained on it so that they can use it in their own role for automating work so it's very helpful not just for our AI engineers and developers when they're doing client work it's it's helpful to teach our entire team so that they know how to create their own AI tools within their work and automate different parts of their own work that they would rather not be doing so it's a win-win across the whole organization when people are using this kind of process so P or PD stands for firstly p is plan which is clearly defining what you need the prompt to do before even writing a single word so this is something that many people chop up on and don't even do D is for draft and this is rapidly generating an initial prompt using an AI tool that we've created that basically rapidly applies the latest research back techniques when it comes to prompting then e is evaluate and this is using a professional prompt engineering software which I'll show you in a second in order to strategically test the prompt that againsts multiple expected inputs so it's rapid testing in a special prompt engineering environment and finally R is for refine which is looking at the responses that you get from the tests we do within software and then making targeted improvements and then repeating that cycle over and over again and what makes this process so powerful is that it combines three critical elements that most people Miss in their prompt engineering which is firstly using research back prompting techniques that are already baked into this process it's using custom AI tools that generate solid first drafts of the prompt for you saving you tons of time and thirdly using a professional prompt engineering IDE that most people have never heard of and the best part is that I'm going to be sharing all of these tools with you for free that we use at morning side so you're going to get access to our prompt planning worksheet our AI prompt generator tool and also a link to the professional prompt IDE that we use as well and so for the rest of this video to show you how this looks in action I'm going to be walking you through each step of the Peter process while actually building a production grade prompt from scratch so you're going to get to see exactly how we do it at morning side with nothing held back so let's start with the first component which is planning the planning phase is where most people go wrong before they even started really they just jump straight into writing a prompt without clearly defining what they needed to do so at morning side we start every prompt with a planning worksheet that answers some very key questions that you need to know before going into things questions like what are the inputs and what data is this prompt actually going to be processing what format does the output need to be in where can I find some high quality examples of input and output pairs which model should be used based on the budget and performance needs of this prompt and of this use case what are the specific requirements or constraints that have come to me from the client or from the the project as a whole so this planning phase typically takes 5 to 10 minutes but it can save hours and hours of frustration later so here on screen we have the prompt planning sheet which you guys are going to be able to get a copy of this is a resource directly out of my accelerator as with some of the other stuff I'm going to be sharing now before we can actually get into planning we need to know what kind of prompt we're going to be writing so in this video I've come up with a hypothetical uh scenario say you are an AI automation agency delivering a project for a client or maybe you're a business and you're doing this internally but we're going to be creating a prompt that is an email order responder um that can basically classify inbound emails to this fantasy uh e-commerce that I've come up with which is called Tech Gear Pro and they sell Tech products online and this prompt that we're about to write is going to be attached to their uh email inbox so every new email is going to be run through the prompt we write and the prompt is basically expected to be able to classify firstly so what kind of query is this is asking about order status are they asking for a return or a refund are they having technical issues and they need technical support or does it fall outside of these so we have four broad categories and that's the classification step and then within each of these we have a bit more detail and a bit more complexity CU I thought that was a bit too simple I want to give you guys a bit more of an advanced example so that you really see how you can apply this to difficult problems within businesses or even within your own maybe you're an employee and you do this stuff so if the email comes in and it classifies it as an order status inquiry if it's been 0 to 7 Days the process is basically to reassure them that hey look usually take 7 to 10 days um please just wait a few days you can track your order here at this ra but if it's over 8 days then this thing is going to go back and say hey look I'm going to be looking into this for you this should be here by now and that's basically buying the human teen time so in the real world um this is an autoresponder that deals with as much of the stuff up front as it can like hey your order should be on the way no worries just check this page but if it's gone over that then it can flag the human team and say hey look um give an instant response to the customer so that they're not sitting there getting all anti because they're not getting a response and basically buys the team time to look into it and then come in with another response the same thing happens with uh return and refund requests here obviously for most e-commerce stores they've got maybe 30-day returns policy so um part of the input data that we're going to be getting with these emails which we'll see in a second this will make a lot more sense in in a second but part of the data that we're going to be getting is also how many days since they Place their order so if it's been less than 30 days since they purchased then the AI autoresponder can automatically give them the process for doing that return and getting the refund if it's been more then they sort of buy a better time for the team and say Hey look let me check if there's anything else I can do um and then the human team can step and same thing for technical support fairly straightforward it's trying to handle most of the situations up front but if it is a b more complex then it's buying time for the human so that's really the the value of the system it gets back to customers as quickly as possible answers as many of those easy to answer questions or just like routing them to the right resources as you can see with these links but in the case where the human does need a step and it's buying time and giving them a response while the humans come and solve it when they come onto their next shift and so this whole thing here is basically a a brief or a write up of what this project would be about and it's got all the context we need in order to be able to build a system to execute on this so now that we know what we're trying to do here we can go back to our planning sheet and fill this out quickly so what are the inputs if any inputs are going to be the uh customer email and the uh number of days since purchase so that's what we're expecting to be passed into this prompt and the sort of system as a whole once it's baked into some kind of workflow automation whereas the IP coming from a any in many cases The Prompt is needing to apply some kind of thinking or some kind of process and in this case the IP is coming from this document here which is all of this uh information so all of this is basically the IP in other cases where you're writing these prompts it may be for you're writing scripting or you're doing short FM content scripting or any kind of other content repurposing in this IP section you could have some Alexi you could have some previous examples of high performing scripts that you put out on your social media Etc but it's basically like what's the juice what's the meat what's actually uh telling it how to transform this data what are the instructions on how it should be doing its task so in this case where's IP coming from is um project brief what format does the output need to be in just a plain text email um respond so the output of this prompt is just going to be the email that the AI has written to respond to so that's not too complicated in a bunch of use cases for this kind of prompt engineering you'd want it to Output in Json so that you can do something with it uh via another step in a workflow automation or you want to extract some information and put into a database so the output format is important and how where can I get two to three high quality examples as you're going to see in the uh for prompt generator tool we do need input output examples in order to get the best possible prompt for my first draft generation in this case I've also included it on uh this write up so I'll give you guys access to all these resources it'll be in my free school Community you can just go onto School join for free it'll take a few minutes to get accepted then in the YouTube resources tab there'll be this video and we'll have all of the resources here including the the perfect prompt generator tool um access to the software that I'm going to be showing you in a second and also uh this document here and the prompt planning sheet Etc so that's all on the school community and here I've generated some example responses so as you can see uh we are expecting inputs of customer email and number of days since purchase so the inputs we have this QA format QA QA question answer question answer this will make a bit more sense when we do the drafting step next but basically we have the input of the email and we have um the number of days since purchase and this is the expected output so we have one pair here and an example input and output and then we have another pair here slightly different and the example output and the input here and another output example as well so these kinds of input and output pairs are really essential to make sure that the output is coming in the right format right tone of voice um and you guys will see in a second how when we bake this into the prompt we can even turn it on and off and see how much of a difference this makes so we have our examples here so we can go back to our prompting sheet um we can say client has provided they are included in the brief and can this be done in one step now in this case I'd say yes but sometimes you may need to chain multiple prompts together for example you may want to throw a whole bunch of information in and get it to write an outline and then it needs to go over that outline and kind of approve that it meets some kind of criteria or way that you like to structure your videos and then you need to do another step after that that's going to actually flesh it out into say a script or or into some piece of content that you want to write so sometimes it's called prompt chaining and you want to do multiple steps where it's you're asking a bit too much for it to do it all in one step but in this case we can do the classification of what type it is and also the generation of the response Allin one go you could split that out into okay let's do a classifier first that's more specialized in the writing of the responses rather than of the classification but in this case I think it's Overkill can this be done in one step yes and which model should be used this really comes down to a number of factors like how much is this going to be used if this is going to be hitting the API maybe thousands of times a month when it comes to the model we need to be thinking about what is this task for what context is it and how many times is it going to be used how difficult is the task so for more difficult tasks you're obviously going to want to go for the more expensive and and smarter models which typically do cost more but at the at the same time if this thing is going to be running thousands and thousands of times a month you may have a bit of a trade-off where hey we need a really intelligent model and we needed a lot so what's the best trade-off here how can I get it to work with the cheapest model possible or if it's only going to be used maybe a handful or a few dozen times a month then you may as well just go for a very smart model and just get the best results from it so in the cases where there's super high volume you're going to want to try and get the cheapest model possible but that can make a little bit more difficult and require a bit more prompt engineering but that's really where the value of of being a good prompt engineer and and having a good process comes from which is you can get better performance out of Cheaper models where other people have to rely on really really expensive ones you can use much cheaper ones with better instructions and get the same or even a better result so in this case because it is a kind of writing task I'm going to go with uh Claude son it 3.7 which is come out recently which is the best smartest writing model but that is really just the starting point we can Benchmark it with that and then we can try to go down to cheaper and cheaper models even within the CLA family but we want to Benchmark it and see just how good we can get it and then if we can use our engineering to work that model down to something cheaper and cheaper and does the client have specific requests none aside from in the brief okay so step two of the planning sheet is actually requirement Gathering so this is prepping everything that we're going to be putting into our AI tool in a second and a lot of this I'm just going to be a to grab directly from our sheet but this is really where you collect a lot of the information that's going to allow you to rapidly draft it in the next step so the purpose of this prompt I can go back to here maybe grab all of this and just take it back over inputs again this is similar to what we had just up here we can just grab these and a bit of context so this is the email and then days of purchase number of days since the cost be used for this is to be used for determining actions um such as refund request and now that we're at this IP section it's asking for the secret Source or methodology which we have in the brief here so if we go response categories and guidelines copy all of this we actually had this information here um maybe I'll just grab that and put it in as well this is the IP or secret source and for the output we can just do similar to what we had before so this step one of exploration is just to get your thoughts out on the page and make sure that you know what you're doing when you head into the second phase here so if we go down and plain text emo response and then for the examples we can go back down and grab our examples from the a client brief well we've got that there and at the bottom of the sheet there's actually another finished example so that if you are doing another prompt you can look at this and get an idea of how you can fill that out for a different kind of prompt and as I said this is from accelerator so there's some other stuff here and the next step is to use our plan in generating our draft which we're going to be covering in a second using our perfect prompt generator tools since we've filled out this entire requirement Gathering section we have completed our planning and now we're ready to move on to the drafting using our AI tool the perfect PR generator so to do the drafting step we need to sign up to relevance AI uh there'll be a link here in the resource so this will be on the school Community um and you can sign up to relevant Ai and once you have you can click this link which is going to allow you to clone a copy of our perfect prom generator into your account and up in the top right hand corner you'll see the Clone button so you want to click this and clone it into your account that means you are going to be responsible for uh paying the billing for it they've got a really good free plan so uh you don't need to worry about that and before we jump into using this and using the plan that we've just written to generate this prompt in a few seconds I just want to go behind the scenes a little bit and you guys will be able to check this out as well once it's in your account but you can see how I've put this tool together so this is what I want you guys to be able to do with the prompts that you create so whether you're an employee whether you're an agency owner whether you're a business owner using relevance AI to create your own tools like this this is an AI tool it's weird because we're using an AI prompt to create an AI tool that creates AI prompts can be a little bit tricky to wrap your head around but this is our perfect prompt generator tool and it's going to take in the use case information which is lot of what we just wrote in the planning and it's also going to take in some input and output examples when we click run on this it's going to send it to a an LM step here where I have baked in all of the latest research such as role prompting uh such as Chain of Thought prompting you have emotion prompting uh use of markdown formatting Etc uh fucha prompting Etc all of the information is already baked into this so what happens when you click on run this tool what it does is it passes in all of the information we did in the plan passes it into this and in a few seconds it applies all of those best practices for prompting to to the uh the inputs and to the use case that you've explained in the inputs so I might actually just change this to uh CL 3.7 and so now you're ready to see just how quickly this uh prompt can be right in the draft so instead of having to do this all manually um you get to do it in a few seconds so if we come back to our planning sheet you go back up what you want to do is take um all of this first section from the purpose down to the examples so we just grab the output go back to relevance Ai and the perfect promp generator tool if we paste this in and that's all of this use case information what is the prompt doing in why give as much information as possible and then we can go down to our input and output examples and we can grab it here go down and if we paste this in here now and click run if we give it a few seconds it's going to write the full draft for this prompt and it's analyzing the use case all the information we provided on it all the rules that we gave it and that IP that we talked about and then it's applying all of the best prompting practices that are research back that we've included in the perfect prompt generator tool just here and now if we scroll down you can see that we've got the full prompt written out so let's just take a scem through it right we have a role you are an expert customer support autor responded for Te Gear Pro R skilled in analyzing customer inquiries identifying to issue types and crafting personalized help with responses that either resolve issues completely or set appropriate expectations for human followup responses are friendly concise R task and the task it seems to have nailed which is analyze the incoming customer email and provide an appropriate response using the step-by-step process so this is Chain of Thought prompting which it's applied carefully read the customer's email identify the order type consider the days since the purchase determine if the issue can be completely resolved Etc so it's taken all of that IP and that juice and instructions that we've provided and it's applied it through a a Chain of Thought method in this case and then here we have things like emotion prompting which is saying we greatly value your ability to strike the perfect balance your expertise is is really really appreciated and this is proven uh through research to improve the performance of the prompts versus when you don't have these emotional words in it as well so it's all woven in there all of the best and latest research and prompt engineering and here we have the inputs that it's going to be expecting and then also as context it's given all of the rules um that the AI needs to be following when responding and it's got our examples in here too um the layout seems to be a little bit off but um it's also got the notes section down the end and what we want to do from here we have technically done the drafting as you can see that's really where we put a lot of effort into the plan because once you throw it into here um and we click this format it then we have it all ready to go nicely formatted and markdown formatting which is another thing that helps with prompt performance and we have it all nicely structured and ready to go into the next step so that is drafting and now we can jump into actually evaluating using our prompt engineering software called prompt metheus okay so here we are this is the software that we use at Morningside for prompt engineering and this has been in sort of secret weapon of mine and the team for a long time I have shared it with my accelerator members but now I think it's time to to probably share uh with you all on the channel here it's what's called a prompt engineering ID e which is an integrated development environment and this is what um coders or developers use when they are writing software so you may be familiar with vs code I'll put some stuff up on screen it may ring a few BS if you've ever watched any kind of tutorials but applications like vs code are basically a a all-in-one development software that allows developers to easily write code manage files uh run code via the terminal it's all all the tools you need really in one place it's the integrated development environment and this is a prompt engineering integrated development environment with all of the best tools that you need to be able to write and test and improve prompts like a like an engineer and taking more of a scientific approach rather than sort of an artsy approach which is what many people do to some success of course I know this is not a sponsored video I'm not getting paid by anyone to promote this or relevance um this is just showing you exactly what we do at morning side so uh the link to sign up to Prometheus will be with all the other links and resources on the school so you can get it there and sign up and once you're inside you'll see something that looks a bit like this and so on the left side here we have our prompts and data sets I'm just going to click add a new prompt here and we can start to build out our prompt within prompt metheus so that we can start to test it and uh evaluate we are in the evaluate part of the Peter process now which means we need to load up our prompt into prompt metheus we need to run some inputs against it and evaluate where we are like how far are we off where we're actually trying to get to is the tone where we need it to be is the prompt actually getting the classification correct so testing all of these things is what the evaluation step is with this initial draft so to get things set up in PR metheus it's actually really really easy if you just go back to relevance Ai and you copy these sections if I copy the roll here and I go back and then I'm going to add this block as roll and you'll see why we start to separate them into these different chunks as it makes it a lot easier for us to test things so we have our roll in here I'm going to add another block here that's the task this time I am going to leave that off I'm just going to skip over the data here for now because we need to do something that's a little bit different for this um which is going to take in the various different inputs we want to test so we can get specifics now now grabbing the context so what you can see me doing here is putting the markdown formatting in so this just helps the AI to understand the structure of the prompt better so this is it's reading this as a heading and this is a like an H2 or or a heading two um so it gives it a lot more structure so this is a specific section this is the task this is the role Etc and here we have a uh a sort of Hing to under the context which is all about the uh the response categories and guidelines as well so now I can do the same with the examples and now we could just grab the notes Here we have all the different components of the prompt split up into the different sections so that's going to make it very easy for us to test in a second but the last step before we actually start evaluating is we need to go and insert the variable or the data set like I said earlier we want to be able to test this against multiple different inputs and determine if it's working as expected so promp neus allows us to do this very easily by going into what's called data sets and if I go back to the uh information that was provided by the client or the brief for the project I have all of the input data here so we have these peers we can go this is the email and then the day since purchases in there as well so if you look at this aside from the from in the subject which you could probably kill and and take out we have the same format in the examples here so the email and then the day since purchase so that's the input format that is expecting and also just to make it a bit easier we can say rename this here with this little rename button on the side you can say this is our um Tech Gear Pro email we can also rename our data dat set here um Tech Gear Pro and so this is one item in the data set this will all make sense in a second so stick with me if we then go add another item we go back to this add another item so I've got 10 of these to add in um these are going to be on the resource you can copy them in if you're following along but the idea is the same that uh you want to have a bunch of different cases that we can test against and I'm just going to sort of go off camera and just add these in now all right so now I have 10 items in my data set and these are dummy emails or dummy data sets that are going to be passed to The Prompt um or hypothetical ones that it should be able to deal with and sort of trying to simulate as much as possible what it would look like in the real world so now we can go back to our uh te Gear Pro email autoresponder prompt and we need to insert into here a data block so we can click on Tech gearpro here and as you can see we can now cycle through all these different cases to test which is super great um and here what we want to do is add on the end if we go back to relevance AI can see that we did have a little bit here and this is basically where it was saying we should input the variables what I like to do is add a little bit more uh we can go like um input data and so this stuff here is obviously going to slot right underneath whatever we have here so we can say classify and write and then we have it now when we run this it's going to fit just under here and this is going to be the uh input data section so I'll click back to one here and we almost almost ready all we need to do now is to select the model so as I said I wanted to run a a clae 3.7 which is the latest one they have gives us the prices here which is super helpful when you are planning these out it has all of the options here for temperature um I might just put this down to 0.1 for temperature and this is something we're going to be testing in the next step and you have things like the token limit um I would probably chop that down to 400 there's things like Json mode so remember when we talked about how you can have different kinds of outputs Json mode allows you to play around with that and and sort of strict and force a Jon formatted output and before we actually run this you will need to set up your API keys so you can come to your account down here by clicking on the bottom left and manage API Keys you can see I've got my anthropic one in there you can go to console. anthropic tocom and you can set up your account um and get your API key and pop it in here you can do it for all these different providers deep seek Gemini grock mistol open AI I've got mine in there too xai and everything so in order to use the models you obviously need to access it through their API but this is how you can manage your API Keys within prom metheus so now we are ready to give this a spin we have all the data in we are going to be testing this first case which is I ordered the premium wireless earbuds on Monday and was wondering when they were arrive I paid for standard shipping and I'm just curious about the timeline so this is someone who purchased recently and is wondering where their product is it should be able to correctly classify this as an email about an order status and then realize that since it's just 4 days since purchase that they should just be sort of prompted to wait a little bit longer so if we give this a spin all right there we go we have our first response and it's saying after analyzing the customer email I can see this so first thing we've noticed already as we test the first one is that it's doing this typical AI thing with the of conversational chat models we have these days it's it's writing a bit of a lead into it so right there that's something that I'd want to tweak and that's going to be in the next step of refining at this point you'd want to be making some notes of the different issues that you want to solve and say hey look I want to add down in the note section what I want to do is add in a a reminder that they should only be replying with the email there should be no other context provided just the email is the output but it also does depend on the system if this was not fully automated and then the AI wasn't actually going to send this email automatically then having this a bit of a summary and sort of Chain of Thought of it explaining why it's doing or why it wrote what it wrote can actually be quite helpful because it increases the accuracy of it but in this case we're assuming that this is going to be for an automated response system and so you wouldn't want it to attack any of that on so it's very important that we prompt that out but for now we can see that it's getting uh the classification and the response right based off our instructions what you can do in PR metheus is grade them very easily because there's the stuff at the top I've given it a thumbs down and this is going to allow me to see over time which ones are good responses of which are this one is a missing order but it's taken longer than 10 days so we should be able to see it's saying hey look let me go and get a human on this there we go it's saying um this delay is unusual I'm personally looking into what happens R I'll keep you updated over the next 24 so it's basically buying time for the human team and again this is probably a thumbs down because it's got the stuff at the start and I think already we can start to see where we want to be tweaking The Prompt and the direction that we want to be making changes so we can jump into the next tip which is going to be the refine stage so in this refining process this is a an iterative thing that you do sort of over and over it goes evaluate and refine and evaluate where you're making changes to the prompt evaluating them again and it's a cycle that you go through until you're happy with the results so in this case as I said we're getting this filler or this uh this Preamble um at the start of all of them which we want to immediately try to buff out um which we can go likely to say the task section here um actually let's go to the specific section I'm going to create a new version so this is one of the tools that we have within prop metheus is that we can add a variant so I'm going to create a variant of these specifics I'm actually going to go back and copy what we had I'm going to put it in here and say add one more to the bottom of this and say your response so I've said your response should not include anything other than the email to be sent to the customer do not include any other prbl reasoning this is because you are part of a an automated resp response system so Le you must not have this extra con content at the start of your responses and then I'll probably try to reinforce this again I might just copy it for now um I will make a duplicate of the context because I think in the uh in the context here we can make it a lot clearer that this is an automated system rather than one that it's allowed to do this kind of stuff at the start so there I've just added in a little bit of extra giving it even an example of the kind of stuff I'm talking about so I've taken a little excerpt out of this and say adding reasoning or any other waffle like XYZ could be disastrous because it would reveal that the system is automated so I have a new version of my context here um I could tack it on to the end of the notes but we have our new version of the specifics with this one added at the end and a new version of the context so now we can give this a run I'll probably run it back over the first two that we just did to make sure that we''ve got them picked off there we go so we're getting the kind of responses we want I think the formatting could be improved but in the real world you'd be building this into some kind of workflow Automation and you could do an additional step after that's a formatter that preps it for say the Gmail or whatever email system that you're using to send and it can make sure that it's actually formatted correctly but in this case uh we've just got the the content correct and the classification correct so now we can give this a nice thumbs up and we can go to the next one and run it again and the goal here is to be able to go through all of these um examples the input examples that I've put into the data set and for it to be categorizing correctly and giving the right kinds of responses based on the examples and based on the uh the guidelines that the customer or the client is given us yep so this one's saying that they're going to be looking into it with than 24 hours that's all good and we can go through to number three here so I think it's pretty clear what the refinement process is from here and the evaluation refinement Loop um you go back and forth until you get in the kind of results that you're looking for I want to give you a few more tips here um just so you can play around more levers that you can pull and there are a few big ones here and this actually really helps you to understand prompt engineering on on a much deeper level because say here we have the switch where we can toggle things on and off so entire sections can be toggled on and off so we can do some interesting experiments like okay if I turn off the roll section and then I run this again is it going to come out much different to what it just did so looks like it's pretty much the same there so it appears that the roll is not actually adding much to it in this case likely because we've got so much context elsewhere but this is a fairly straightforward use case and we are using quite a powerful model which is something we can test in a second actually is bringing that model quality down and seeing if we can get the same results with an even cheaper model um or another good one that I like to do is turning off the examples and seeing if they're actually making a difference because in a lot of cases your examples are going to be a massive massive amount wall of text um especially when you have large inputs um and you maybe have a a whole V YouTube video that you want to put in and the transcript is the input and and therefore your example section is gigantic um but in this case let's just turn off our examples and we'll give it a spin and see how this one comes out I actually there is a bit of a difference see here we have the roll and the task on and it gave us this nice nicely formatted exactly how we wanted it and then if we go up we have taken the roll out and one we getting the subject in there so it seems to be a lot less accurate it's not following the exact response format that we've asked for here which is it's obviously ignoring the fact that we said look you you can't put in any other stuff that's going to reveal that this is not a respond so this would be a fail this one would be good and here you can see we've added the roll back in but taken out the examples and we're seeing similar issues again so the role in this case actually appears to be doing quite a lot it's paying a lot more attention to the instructions we give it and also understanding that it's part of an autoresponder system and therefore this stuff can't be included it appears that the example section where we have it here but not in here is not actually doing that much so you can see how you can start to test and figure out what parts of the prompt are actually moving the needle what parts are needed um that actually isn't isn't a green thumb so if we turn everything back on I've shown you guys how you can create new versions of sections easiest way is to right click on one of these tabs and click duplicate and it's going to give you a and a duplicate version of that you can make changes to it chops and bits out and the last test that I'd want to do here is see if we can get the same performance um let's just run through a couple of these there we go um four five and six so we'll just get these filled out using the uh the best prompt that we know is is performing how we want it to and there we go we have all of these responses let's just double check this last one number six this is a technical support one and it seems to be dealing with that correctly there so that's good this is performing as expected now if we change this down to a cheaper version of claude's model so 3.5 hiu and we run this on six here and then we compare them now actually this probably isn't a fair comparison because uh this part with this tech support kind of questions it is coming up with its own AI generated responses to we don't actually have a knowledge base on tech support which is what you do with more of an agent based build but if we go back to um let's just do one two and three again so an easy way of doing this would be to just go to High and we'll run one and then we'll switch to Sonet 3.7 and we'll compare it side by side and it appears that we are getting yeah pretty much the same thing and you can see the cost has gone from 83 cents here down to 24 cents it's taken almost a second less to get this same kind of result so prop metheus is great for giving you some helpful outputs like this obviously this is uh 3 times cheaper um than this one and if we go for back to uh 3.5 here and we test number two and then we go to uh 3.7 we test number two again you can see okay we are getting yep pretty much the same response um exactly the same on both of them there's also other things you can play with if you're doing more writing based tasks like come into the temperature and play around with the temperature but I think I've given you a pretty good rundown of the entire process of uh how we build prompts it's Peter process we write the plan out use that plan put it into the tool to rapidly generate the draft fit all the pieces into promt metheus here set up your data sets and then rapidly go through and test all of them to make sure it's working and that refinement cycle and the different levers you have to pull on like turning sections off creating new uh variations of certain sections playing around with the model the temperature etc those are all your tools and I want to just give you a final sort of Keystone or cap to this entire project which is how you can turn these prompts into uh usable tools that can actually be really helpful for you either as an employee or or or business owner is to actually go back to relevant Ai and to make handy internal tools for yourself or for other people in the company or even for clients if you're an AI Agency on relevance here say if you wanted to turn this into a autoresponder tool not a super great example but you'd probably do this more for say like uh content generation where you put in a couple of inputs and it gives you uh through your prompt engineering work it's going to be a to transform that into an ideal kind of content for whatever platform you're choosing but in this case we can then take all of these um this is of course taking in two inputs so we could say input one is going to be email content we can go a have be short text and we can go um days since per then we can go to an llm step here and what we can do is just start to build out our prompt again inside relevant Ai and then we can just pop these in here and go email content and put in the email there and if we go at days since purchase don't need to put this in of course cuz we're going to be passing that in within relevance and we can just dump the rest of the prompt there here now that we've finalized it and really worked it through so the thing is that testing it within something like relevance or any other kind of workflow Builder would be a lot more difficult and take a whole lot more time and you don't have anywhere near as much control um and it's just really an inferior way of testing and building the prompt so we can then replicate The Prompt that we chose here which is going to be our claw 3.5 hiu and if we save and publish we now have our own email reply generator tool so say if you want to make a tool like this use your your pro engineering skills or get your your staff or team to do this and then you can go to use how you could even go share this is usually how I'd like to share them around the company copy this and now you've got this cool Standalone tool where anyone in the company can access this tool and I can go back and do the exact same thing we did and we put in our where's my order and we've got four days since purchase and Bam that's how you can create your own handy dandy AI tools um and use them sort of wherever you need within your business within your personal life um and using that prompt Engineering Process and there you go we're getting the kind of outputs we want and this can be used by API all sorts of different things but that's the prompt Engineering Process first and foremost that I wanted to teach you and while that planning phase can feel a bit clunky at first it's so good if you just rip through that and then you paste that into the the AI tool to generate it and before you know it you've got a like 90% of the way there with a few tweaks needed as you saw but being able to extract more value out of these AI models and not having pay for that tradeoff of okay I don't have the time to spend to make a really good prompt this is the the whole system that we use at warning sides so I hope that's been helpful for you guys if you want to get all the resources and the perfect prompt generator tool I will even give you access to our perfect prompt generator tool for AI agents which is a bit more complex but I know a lot of you guys are building agents now and we have a really really good tool for doing that as well so that's all going to be linked in the first link in the description will be to our school you go to YouTube resources on school once you're inside and you'll find this video and all the resources will be on there also my entire free course as well for building AI businesses building AI systems building agents and everything is in there as well so that's all for the video guys if you want to check out my recent video which is massive breakdown of how to build AI agents like probably the best resource on the entire internet right now about how to learn AI agent building um if you're a non-technical person or even if you are a technical person that's up there but that's all for the video guys thank you so much for watching and I will see you in the next one